

Stiefel 流形上的梯度算法及其在特征提取中的应用

章建军^{*①} 曹杰^② 王源源^①

^①(南京航空航天大学电子信息工程学院 南京 210016)

^②(南京航空航天大学无人机研究院 南京 210016)

摘要: 为了提高系统特征提取算法的计算效率、减少占用的存储空间和简化程序设计, 该文基于 Riemann 流形上优化算法的几何框架, 提出了改进的 Stiefel 流形上的梯度下降算法。根据不同要求采用不同的测地线计算公式, 并使用多项式逼近测地线方程, 同时采用了秦九韶-Horner 多项式算法及线搜索、变步长的方法。以主分量分析问题为例, 详细讨论了 Stiefel 流形上的梯度算法在其中的应用。理论分析和实验结果均表明, 此方法可以在确保迭代矩阵列向量单位正交性的同时获得更好的计算效率和收敛速度, 并且更容易实现。

关键词: Stiefel 流形; 特征提取; 梯度算法; 测地线; 主分量分析

中图分类号: TP391

文献标识码: A

文章编号: 2095-283X(2013)03-0309-05

DOI: 10.3724/SP.J.1300.2013.13048

Gradient Algorithm on Stiefel Manifold and Application in Feature Extraction

Zhang Jian-jun^① Cao Jie^② Wang Yuan-yuan^①

^①(College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)

^②(Unmanned Aerial Vehicle Research Department, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)

Abstract: To improve the computational efficiency of system feature extraction, reduce the occupied memory space, and simplify the program design, a modified gradient descent method on Stiefel manifold is proposed based on the optimization algorithm of geometry frame on the Riemann manifold. Different geodesic calculation formulas are used for different scenarios. A polynomial is also used to lie close to the geodesic equations. JiuZhaoQin-Horner polynomial algorithm and the strategies of line-searching technique and change of the step size of iteration are also adopted. The gradient descent algorithm on Stiefel manifold applied in Principal Component Analysis (PCA) is discussed in detail as an example of system feature extraction. Theoretical analysis and simulation experiments show that the new method can achieve superior performance in both the convergence rate and calculation efficiency while ensuring the unitary column orthogonality. In addition, it is easier to implement by software or hardware.

Key words: Stiefel manifold; Feature extraction; Gradient algorithm; Geodesic; Principal component analysis

1 引言

系统特征信息分析方法包括主分量分析(Principal Component Analysis, PCA)、次分量分析(Minor Component Analysis, MCA)和独立分量分析(Independent Component Analysis, ICA)^[1]等, 其广泛应用于数据压缩、波达方向估计、信号滤波等信号处理领域^[2,3]。近十年来, 众多专家学者提出

了各种算法^[4,5]。然而, 现行的各种算法大都是基于欧氏空间的, 这些学习算法并没有充分利用约束集为 Stiefel 流形这一事实, 因而也就不能充分利用 Stiefel 流形的几何性质。另外, 这类算法往往需要进行特征分解, 而特征分解运算量大, 算法复杂, 工程难以实现。

本文首先详细讨论了 Riemann 流形上梯度算法, 重点讨论了 Stiefel 流形上的梯度算法。其次, 对测地线公式作了深入分析与比较, 为了能够在工程实际中应用, 作了较大改进, 提出了改进的梯度下降算法 M-GDM(Modified Gradient Descent

2013-05-22 收到, 2013-08-30 改回; 2013-09-03 网络优先出版

国家自然科学基金(61106018)资助课题

*通信作者: 章建军 hlyup@163.com

Method)。最后,以主分量分析问题为例,详细讨论如何在 Stiefel 流形上应用梯度下降算法。

2 Riemann 流形上的梯度算法

作为求解非线性最优化问题的新方法,基于流形的几何优化算法核心思想是把约束集视为高维空间中的曲面,也就是(微分)流形^[6]。当在流形上指定了一个 Riemann 度量(也就是在流形的每一点的切空间上指定一个对称、正定的 2 阶协变张量)后,成为 Riemann 流形。

测地线^[7]是 Riemann 几何中非常重要的概念,定义如下:

定义 1^[7] 设 M 为 Riemann 流形,局部坐标为 $\boldsymbol{x} = (x^1, x^2, \dots, x^d)$, 流形上光滑曲线定义为 $\gamma = [a, b] \rightarrow M$, 记 $x^i(t) = x^i(\gamma(t))$, Γ_{jk}^i 为 Christoffel 符号, 而且满足下面的 2 阶微分方程组

$$\frac{d^2 x^i(t)}{dt^2} + \Gamma_{jk}^i(\boldsymbol{x}(t)) \frac{dx^j(t)}{dt} \frac{dx^k(t)}{dt} = 0, \quad \forall i = 1, 2, \dots, d \quad (1)$$

则光滑曲线 γ 称作流形上的测地线。式(1)采用了 Einstein 求和约定,即同一指标在表达式中恰好出现两次,则关于此指标求和。测地线又称为短程线,指 Riemann 流形上连接两点之间的最短曲线。比如,球面上两点之间的测地线就是连接两点的大圆。在选定局部坐标系后,求解式(1)对应的 2 阶微分方程组就可以求得测地线方程,一般而言,直接求解式(1)是非常困难的。

一方面,由于 Riemann 流形本质上是一种弯曲的空间,所以迭代估计应该沿着流形的测地线进行^[8];另一方面, n 维 Euclidean 空间是一种曲率处处恒为 0 的特殊的 Riemann 流形,所以无约束最优化和约束最优化都可以统一成 Riemann 流形上的最优化。然而由于 Euclidean 空间中的直线就是测地线,所以无约束优化的搜索方向可以沿着空间中任何方向,而约束最优化的搜索方向则应沿着 Riemann 流形上测地线所对应的切方向。据此,可以解决梯度算法中的一个关键问题——搜索方向,并可以得出 Riemann 流形上无约束优化和约束优化的梯度类算法的统一框架^[9]。

(1) 初始化: 随机选取流形上的初始点 x_0 。

(2) 计算负梯度方向对应的测地线。因为 Euclidean 空间中测地线就是直线,所以对于无约束最优化,就是 $-\frac{\partial f}{\partial \boldsymbol{x}} \Big|_{\boldsymbol{x}=x_{k-1}}$ 。然而对于约束最优化,则应先计算 $-\frac{\partial f}{\partial \boldsymbol{x}}$, 然后把 $-\frac{\partial f}{\partial \boldsymbol{x}}$ 投影到流形在点 x_{k-1} 处的切空间上。

(3) 沿测地线移动一定的步长,得到新的迭代点 x_k 。

(4) 若点 x_k 满足收敛准则,则最优解为 x_k 。否则返回步骤(2),用 x_k 作为新的起始点,继续进行迭代。

在一定条件下,可以证明上面的梯度算法是收敛的。证明过程较长,可以参阅文献[8]。

3 Stiefel 流形及其测地线方程

Stiefel 流形定义为全体 $n \times r$ 半正交矩阵的集合^[10],用符号 $\text{St}(n, p)$ 表示,即

$$\text{St}(n, p) = \{ \boldsymbol{W} \mid \boldsymbol{W}^T \boldsymbol{W} = \boldsymbol{I}_p, \boldsymbol{W} \in R^{n \times p}, n \geq p \} \quad (2)$$

值得指出的是,Stiefel 流形是紧致的,即参数的取值范围是一个有限闭集。关于这一点,可以简单的说明一下,设有 $\boldsymbol{W} \in \text{St}(n, p)$, 则有

$$\left. \begin{aligned} \boldsymbol{w}_i^T \boldsymbol{w}_i &= 1, \quad 1 \leq i \leq p \\ \boldsymbol{w}_i^T \boldsymbol{w}_j &= 0, \quad 1 \leq i, j \leq p, i \neq j \end{aligned} \right\} \quad (3)$$

所以有 $\sum_{j=1}^n |w_{i,j}|^2 = 1$, 也即有 $|w_{i,j}| \leq 1$, 即元素的取值在一个有限的范围内。注意到这一点,后面关于测地线的计算公式可以进一步近似并简化。

Riemann 流形上梯度下降算法的另一个核心问题测地线的计算。在微分流形与 Riemann 几何中测地线方程的推导往往采用内蕴的方式,测地线对应一组微分方程组的解,直接求解并应用到实际中有很大的困难。然而 Stiefel 流形上的测地线却有相对简单的表达式,文献[10]给出了 Stiefel 流形上测地线的一种计算公式,以定理表述如下:

定理 1^[10] 设 \boldsymbol{Y} 和 \boldsymbol{H} 均为 $n \times p$ 的矩阵,满足 $\boldsymbol{Y}^T \boldsymbol{Y} = \boldsymbol{I}_p$, $\boldsymbol{A} = \boldsymbol{Y}^T \boldsymbol{H}$ 为反对称矩阵,则在 Stiefel 流形上 \boldsymbol{Y} 点处沿 \boldsymbol{H} 方向的测地线为:

$$\boldsymbol{Y}(t) = \boldsymbol{Y} \boldsymbol{M}(t) + \boldsymbol{Q} \boldsymbol{N}(t) \quad (4)$$

其中, $\boldsymbol{Q}_{n \times p}$, $\boldsymbol{R}_{p \times p}$ 为 \boldsymbol{K} 矩阵的 QR 分解,而 $\boldsymbol{K} = (\boldsymbol{I} - \boldsymbol{Y} \boldsymbol{Y}^T) \boldsymbol{H} = \boldsymbol{Q} \boldsymbol{R}$, $\boldsymbol{M}(t)$ 和 $\boldsymbol{N}(t)$ 是由矩阵指数

$$\begin{bmatrix} \boldsymbol{M}(t) \\ \boldsymbol{N}(t) \end{bmatrix} = \exp \begin{bmatrix} \boldsymbol{A} & -\boldsymbol{R}^T \\ \boldsymbol{R} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{I}_p \\ \mathbf{0} \end{bmatrix} \quad (5)$$

确定的 $p \times p$ 的矩阵。

当 p 较小时,上述方法的计算量相对较小。然而当 p 接近 $n/2$ 时,上述方法的计算量就很大了,而且占用的存储空间大、程序设计复杂。文献[11]给出了 Stiefel 流形上测地线新的计算公式,使得当 p 接近 $n/2$ 或 $2p > n$ 时(比如在独立分量分析 ICA 中),计算量变得更小。新的测地线方程为:

$$\left. \begin{aligned} \varphi_{\text{St}(n,p)}(\mathbf{W}, \mathbf{V}, t) &= \exp(t(\mathbf{D}\mathbf{W}^T - \mathbf{W}\mathbf{D}^T))\mathbf{W} \\ \mathbf{D} &= \left(\mathbf{I} - \frac{1}{2}\mathbf{W}\mathbf{W}^T \right) \mathbf{V} \end{aligned} \right\} \quad (6)$$

上式表示起点为 \mathbf{W} , 方向为 \mathbf{V} 的测地线方程, 即 $\gamma: [t_1, t_2] \rightarrow M$ (M 为 Stiefel 流形), 并且满足

$$\varphi_M(\mathbf{W}, \mathbf{V}, t) = \gamma(t), \quad \gamma(0) = \mathbf{W}, \quad \left. \frac{d\gamma}{dt} \right|_{t=0} = \mathbf{V} \quad (7)$$

文献[11]同时也给出了相应的测地线逼近公式, 但是需要矩阵求逆和矩阵乘方。

测地线计算公式(4)需要 QR 分解及矩阵指数运算, 测地线计算公式(6)需要矩阵求逆及矩阵乘方运算, 这两种方法不仅计算量大, 而且占用较大的存储空间、编程复杂。为了减小计算量, 同时简化程序设计, 必须简化测地线的计算方法(当然必须尽可能地满足单位列正交性的约束条件, 这是在流形上进行优化的根本目的)。

鉴于此, 本文对上面两种算法中的矩阵指数运算作近似。考虑矩阵指数函数的 Taylor 展开式:

$$\begin{aligned} \exp(t\mathbf{X}) &= \mathbf{I} + t\mathbf{X} + \frac{1}{2!}(t\mathbf{X})^2 + \frac{1}{3!}(t\mathbf{X})^3 \\ &+ \dots + \frac{1}{n!}(t\mathbf{X})^n + \dots \\ &= \sum_{n=0}^{\infty} \frac{1}{n!}(t\mathbf{X})^n \end{aligned} \quad (8)$$

由于流形上指数映射往往限于局部, t 取值往往比较小(比如取 0.01), 又由于 Stiefel 流形为紧致的, 所以最终切向量 \mathbf{X} 中的元素取值较小, t 和 \mathbf{X} 中各个元素相乘后, 再进行矩阵乘法运算, 最终的矩阵元素取值就更小, 略去这些项, 误差会非常小, 对正交性约束影响也自然非常小。

以上定性分析了误差较小的原因, 下面进行定量分析, 给出误差一个较好的上界。设截取 n 项, $\|\cdot\|$ 表示矩阵范数, 并令 $\|\mathbf{X}\| = 1$, 则有

$$\begin{aligned} \varepsilon &= \left\| \sum_{i=n}^{\infty} \frac{1}{i!}(t\mathbf{X})^i \right\| \leq \sum_{i=n}^{\infty} \left\| \frac{1}{i!}(t\mathbf{X})^i \right\| \\ &\leq \sum_{i=n}^{\infty} \frac{1}{i!} t^i \|\mathbf{X}\|^i = \sum_{i=n}^{\infty} \frac{1}{i!} t^i \end{aligned} \quad (9)$$

由 Stirling 公式 $n! = \sqrt{2\pi n}(n/e)^n e^{\theta/12n}$ ($0 < \theta < 1$), 得 $n! > \sqrt{2\pi n}(n/e)^n$ 。所以有

$$\begin{aligned} \varepsilon &\leq \sum_{i=n}^{\infty} \frac{t^i}{i!} < \sum_{i=n}^{\infty} \frac{t^i}{\sqrt{2\pi i}(i/e)^i} = \sum_{i=n}^{\infty} \frac{(te)^i}{\sqrt{2\pi i}(i)^i} \\ &< \frac{1}{\sqrt{2\pi n}} \sum_{i=n}^{\infty} \frac{(te)^i}{n^i} = \frac{1}{\sqrt{2\pi n}} \frac{(te)^n}{n^n} \frac{n}{n-te} \end{aligned} \quad (10)$$

当 n 较大, t 较小时, 式(10)还可以作进一步近似

$$\varepsilon < \frac{1}{\sqrt{2\pi n}} \frac{(te)^n}{n^n} \frac{n}{n-te} \approx \frac{1}{\sqrt{2\pi n}} \frac{(te)^n}{n^n} \quad (11)$$

取 $n=5, t=0.01$, 由式(10)得 $\varepsilon < 8.52 \times 10^{-13}$, 由式(11)得 $\varepsilon < 8.47 \times 10^{-13}$, 而由式(9)得 $\varepsilon < 8.35 \times 10^{-13}$ 。可以看出, 式(10)和式(11)都有较好的估计精度, 给出了一个较好的误差上界。由此误差上界, 结合数据的动态范围, 也可以较好地确定截取的项数。

为了进一步减少计算量, 对多项式运算采用秦九韶-Horner 算法^[12], 即:

$$\begin{aligned} \text{设多项式为 } p_n(\mathbf{X}) &= \alpha_n \mathbf{X}^n + \alpha_{n-1} \mathbf{X}^{n-1} + \alpha_{n-2} \mathbf{X}^{n-2} + \dots + \alpha_1 \mathbf{X} + \alpha_0 \mathbf{I}, \text{ 改变计算顺序为:} \\ p_n(\mathbf{X}) &= (\dots((\alpha_n \mathbf{X} + \alpha_{n-1} \mathbf{I})\mathbf{X} + \alpha_{n-2} \mathbf{I})\mathbf{X} + \dots \\ &+ \alpha_1 \mathbf{I})\mathbf{X} + \alpha_0 \mathbf{I} \end{aligned} \quad (12)$$

此时仅仅需要 n 次矩阵乘法和 n 次矩阵加法。

梯度算法的第 3 个关键问题是迭代步长的选取。迭代步长的选取本质上是一种“线搜索”或“1 维搜索”技术, 为了使算法具有更好的性能, 应区分批处理和在线处理这两种不同的情形, 并采用不同的策略, 而以往大多数算法都忽略了这一重要问题。

对于批处理情形, 此时相关矩阵已经比较好地估计出来(用时间平均取代统计平均), 一定程度上, 此时已经是确定性的最优化问题, 所以采用不精确线搜索技术——Armijo 准则^[13]。欧氏空间中, Armijo 准则是指对于给定 $\beta \in (0, 1), \sigma \in (0, 0.5)$, 令步长因子 $\alpha_k = \beta^{m_k}$, 其中 m_k 为满足下列不等式的最小非负整数:

$$f(x_k + \beta^m d_k) \leq f(x_k) + \sigma \beta^m (\nabla f(x_k))^T d_k \quad (13)$$

其中 f 为目标函数, d_k 为下降方向, ∇f 为目标函数梯度。Riemann 流形上的 Armijo 准则及收敛性分析见文献[13]。

在一些信号处理领域, 为了能够实时处理, 必须采用在线处理的方式, 由于此时相关矩阵的估计仅仅采用当前时刻的采样值, 所以相关矩阵是非常不精确的和随机的。为了减小稳态误差、提高性能, 应采用变步长的方法。变步长的方法较多, 本文采用比较简单的变步长公式 $\eta(n) = \exp(-n \cdot 0.05 - 1)$, 其它的变步长方法可以参考自适应滤波相关的文献^[14,15]等。

4 Stiefel 流形上的特征提取

PCA, SCA 和 ICA 等问题最终归结为求解某一矩阵 \mathbf{W} , 而求解矩阵 \mathbf{W} 是通过最小化 Stiefel 流形上某一代价函数实现的, 即转化为非线性约束最优化问题, 然后通过梯度下降算法求得 \mathbf{W} 。设代价函数为 $J(\mathbf{W})$, 相应的梯度为 $\nabla J(\mathbf{W})$ 。值得注意的是, 此时仍然是在欧氏空间中考虑问题(未考虑约束条件), 按

Riemann 流形上梯度算法的一般原则必须把梯度向量 $\nabla J(\mathbf{W})$ 投影到流形上点 \mathbf{W} 处的切空间上。对于 Stiefel 流形, 有下面结论, 以定理形式表述为:

定理 2^[8] 设 Stiefel 流形上 1 阶连续可微函数为 $f(x)$, x 为 Stiefel 流形上的点。当把 Stiefel 流形嵌入到 Euclidean 空间后, 在 Euclidean 空间中的梯度为 $\nabla f = (\partial f / \partial x_{i,j})$, 投影 $\nabla f = (\partial f / \partial x_{i,j})$ 到流形上点 x 处的切空间得 $\tilde{\nabla} f = (\partial f / \partial x_{i,j})$, 则 ∇f 和 $\tilde{\nabla} f$ 有关系: $\tilde{\nabla} f = \nabla f - x(\nabla f)^T x$ 。

据此, 由 Riemann 流形上梯度算法的一般框架, 可以得出 Stiefel 流形上梯度算法的一般步骤:

- (1) 确定待求矩阵 \mathbf{W} 对应的代价函数 $J(\mathbf{W})$, 并选择初始点 \mathbf{W}_0 。
- (2) 求代价函数 $J(\mathbf{W}_k)$ 的梯度 $\nabla J(\mathbf{W}_k)$, 并按定理 2 把 $\nabla J(\mathbf{W}_k)$ 投影到流形上点 \mathbf{W}_k 处的切空间上, 设为 $\tilde{\nabla} J(\mathbf{W}_k)$, $\tilde{\nabla} J(\mathbf{W}_k)$ 就是相应的搜索方向。
- (3) 使用“线搜索”技术确定步长参数 α , 并沿测地线移动步长 α , 得到新的迭代点 \mathbf{W}_{k+1} 。
- (4) 若点 \mathbf{W}_{k+1} 满足收敛准则, 则最优解为 \mathbf{W}_{k+1} 。否则用 \mathbf{W}_{k+1} 作为新的起始点, 继续进行迭代, 直到收敛。

必须指出的是, 测地线计算公式的选择很重要, 当待求的主分量数目较少时, 应选用式(4), 否则应选用式(6)。

下面以 PCA 问题为例, 讨论如何在 Stiefel 流形上应用梯度下降算法。PCA 问题可以等价于 Stiefel 流形上非线性约束最优化问题:

$$\max J(\mathbf{W}) = \frac{1}{2} \text{tr}(\mathbf{W}^T \mathbf{R} \mathbf{W}), \quad \text{s.t. } \mathbf{W}^T \mathbf{W} = \mathbf{I} \quad (14)$$

其中, \mathbf{R} 为输入随机向量 \mathbf{x} 的协方差矩阵, 即 $\mathbf{R} = \mathbf{E}(\mathbf{x}\mathbf{x}^T)$ 。

目标函数 $J(\mathbf{W})$ 在 Euclidean 空间中相应的梯度为 $\nabla J(\mathbf{W}) = \mathbf{R}\mathbf{W}$, 由定理 2, 可以很容易得到 Stiefel 流形上相应的梯度为(注意到 \mathbf{R} 为对称矩阵):

$$\tilde{\nabla} J(\mathbf{W}) = \mathbf{R}\mathbf{W} - \mathbf{W}\mathbf{W}^T \mathbf{R}\mathbf{W} \quad (15)$$

在线处理时 \mathbf{R} 用瞬时估计值代替, 即 $\mathbf{R} = \mathbf{x}(n)\mathbf{x}^T(n)$ 。按 Riemann 流形上梯度下降算法, 得到相应的随机梯度算法的迭代公式:

$$\mathbf{W}(n+1) = \exp_{\mathbf{W}(n)} \left[\eta(n) (\mathbf{x}(n)\mathbf{x}^T(n)\mathbf{W}(n) - \mathbf{W}(n)\mathbf{W}^T(n)\mathbf{x}(n)\mathbf{x}^T(n)\mathbf{W}(n)) \right] \quad (16)$$

式(16)中, $\exp_{\mathbf{W}}(\mathbf{v})$ 为 Riemann 流形上的指数映射, 表示流形上过点 \mathbf{W} , 切方向为 \mathbf{v} 的测地线, $\mathbf{x}(n)$ 为 n 时刻的采样值, $\eta(n)$ 为步长参数。

至于 MCA 和 ICA 等问题, 只要知道代价函数^[1], 很容易依次类推。

5 数值仿真与结论

记文献[11,16]提出的算法为 E-GDM(Exponential Gradient Descent Method)。

实验 1 比较正交性。求 5 维随机向量的第 1, 第 2 主分量, 分别设为 \mathbf{w}_1 和 \mathbf{w}_2 。理论上, \mathbf{w}_1 和 \mathbf{w}_2 应相互正交, 即 $\mathbf{w}_1^T \mathbf{w}_2 = 0$ 。但是, 采用多项式近似后, 必然有误差, 即 $\mathbf{w}_1^T \mathbf{w}_2 \neq 0$, 直接取绝对值作为正交性误差的度量, 即 $|\mathbf{w}_1^T \mathbf{w}_2|$ 。实验时随机选取的 5 维随机向量的协方差矩阵为:

$$\mathbf{R} = \begin{bmatrix} 0.3722 & 0.7712 & 0.0569 & 0.2335 & 0.3395 \\ 0.7712 & 7.5237 & 1.7506 & 2.9652 & -2.4571 \\ 0.0569 & 1.7506 & 3.9186 & 1.0447 & 0.4627 \\ 0.2335 & 2.9652 & 1.0447 & 1.6857 & -0.4566 \\ 0.3395 & -2.4571 & 0.4627 & -0.4556 & 17.4008 \end{bmatrix}$$

多项式取 7 项, 步长参数取 0.01, 迭代 100 次, 结果如图 1。

从实验可以看出, 用多项式近似和使用指数映射(exp)其性能是非常接近的, 对正交性约束的影响极小, 可以认为正交性得以保持。

实验 2 比较运行时间。由于 matlab 中 QR 分解算法是用内嵌汇编实现的, 为了便于比较, 故使用 C 语言编程实现两种算法。多项式取 7 项, 向量维数从 40 取到 100, 每次取 10000 个采样数据, E-GDM 算法与 M-GDM 算法占用的时间比如图 2。

可以看出时间比基本稳定在 15 左右, 然而本文的算法简单、占用存储空间也小。

实验 3 比较收敛性。求解第 1 主分量, 误差记为 $\|\mathbf{w} - \hat{\mathbf{w}}\|^2$, 其中 \mathbf{w} 为精确值, $\hat{\mathbf{w}}$ 为每次迭代的估计值。多项式取 7 项, 迭代 100 次, 实验结果如图 3。

从实验可以看出, 由于改进的算法采用了“线搜索”技术, 收敛速度更快, 不仅迭代步数变小, 而且误差也更小。

6 结束语

如文中所指出, 如果在高维数据(设为 n)中仅仅提取少量主分量(设为 p), 则应选用测地线计算公式(4); 当提取的主分量数接近 $n/2$ 或者当 $p > n/2$ 时, 则应选用第 2 个测地线计算公式(6)。在独立分量分析问题中, 几乎总是使用计算公式(6)。总而言之, 应具体问题具体对待。

在 Riemann 流形(包括 Stiefel 流形)上发展其它类型的优化算法, 并将其应用到信号处理中, 以提高算法的效率, 不仅是重要的思路, 也是一个值得研究的方向。

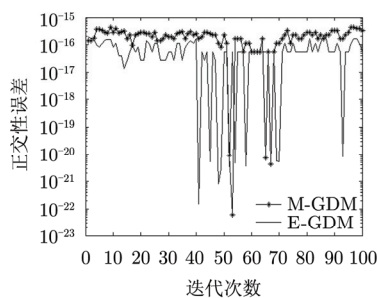


图1 正交性误差比较

Fig. 1 The comparison of orthogonality error

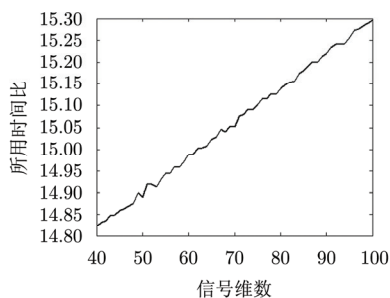


图2 E-GDM 和 M-GDM 占用的时间比

Fig. 2 The ratio of occupying time between E-GDM and M-GDM

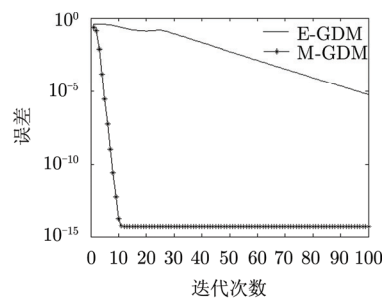


图3 收敛速度性能曲线

Fig. 3 The performance curve of convergence rate

参考文献

- [1] Kong X Y, Hu C H, and Han C Z. A dual purpose principal and minor subspace gradient flow[J]. *IEEE Transactions on Signal Processing*, 2012, 60(1): 197-201.
- [2] Eldar Y C and Oppenheim A V. MMSE whitening and subspace whitening[J]. *IEEE Transactions on Inform Theory*, 2003, 49(7): 1846-1851.
- [3] Chang D X, Feng D Z, and Zheng W X. A fast recursive total least squares algorithm for adaptive IIR filtering[J]. *IEEE Transactions on Signal Processing*, 2005, 53(3): 957-965.
- [4] Ouyang S and Bao Z. Fast principal component extraction by a weighted information criterion[J]. *IEEE Transactions on Signal Processing*, 2002, 50(8): 1994-2002.
- [5] Moller R and Konies A. Coupled principal component analysis[J]. *IEEE Transactions on Neural Networks*, 2004, 15(1): 214-222.
- [6] Rapcsak T. On minimization on Stiefel manifold[J]. *European Journal of Operational Research*, 2002, 143(2): 365-376.
- [7] Jurgen J. *Riemannian Geometry and Geometric Analysis*[M]. Berlin Heidelberg, NY, US, Springer, 2005: 13-20.
- [8] Ferrira O P and Oliveira P R. Sub gradient algorithm on Riemannian manifolds[J]. *Journal of Optimization Theory and Application*, 1998, 97(1): 93-104.
- [9] Yang Y. Global convergent optimization algorithm on Riemannian manifold: uniform framework for unconstrained and constrained optimization[J]. *Journal of Optimization Theory and Application*, 2007, 132(2): 245-265.
- [10] Edelman A, Arias T, and Smith S. The geometry of algorithm with orthogonality constraints[J]. *SIAM Journal Matrix Analysis and Applications*, 1998, 20(2): 302-353.
- [11] Nishimor Y and Akaho S. Learning algorithm utilizing quasi-geodesic flows on the Stiefel manifold[J]. *Neurocomputing*, 2005, 67: 106-135.
- [12] 徐树方, 钱江. 矩阵计算六讲[M]. 北京: 高等教育出版社, 2011: 70-71.
Xu Shu-fang and Qian Jiang. *Six Lectures on Matrix Computation*[M]. Beijing: Higher Education Press, 2011: 70-71.
- [13] 黄建国, 孙连山, 叶中行. 黎曼流形上带 Armijo 准则步长准则优化算法[J]. 上海交通大学学报, 2002, 36(2): 267-271.
Huang J G, Sun L S, and Ye Z X. Optimization algorithm with Armijo rule on Riemann manifold[J]. *Journal of Shanghai Jiao Tong University*, 2002, 36(2): 267-271.
- [14] 张贤达. 现代信号处理[M]. 北京: 清华大学出版社, 2012: 199-200.
Zhang X D. *Modern Signal Processing*[M]. Beijing: Tsinghua University Press, 2012: 199-200.
- [15] Yang H H. Series updating rule for blind separation derived from scoring[J]. *IEEE Transactions on Signal Processing*, 1999, 47: 2279-2285.
- [16] 段玲, 黄建国. 主成分分析的一个黎曼几何随机算法[J]. 上海交通大学学报, 2004, 38(1): 71-74.
Duan L and Huang J G. A Riemannian geometry underlying stochastic algorithm for adaptive principal component analysis[J]. *Journal of Shanghai Jiao Tong University*, 2004, 38(1): 71-74.

作者简介

章建军(1988-), 男, 江苏南京, 硕士研究生, 主要研究方向为优化算法与盲信号处理。

E-mail: hblyup@163.com

曹杰(1963-), 男, 研究员, 研究方向为信号处理、数字图像处理。

王源源(1988-), 女, 江苏镇江, 硕士研究生, 主要研究方向为数据压缩。